AD-A068 146 TEXAS A AND M UNIV COLLEGE STATION INST OF STATISTICS F/G 12/1

AN ANALYSIS OF THE WEIBULL FAMILY OF DISTRIBUTIONS FROM A QUANT--ETC(U)

FEB 79 J M WHITE DAAG29-78-G-0180

UNCLASSIFIED TR-A-3 ARO-16228.4-M NL

| OF |

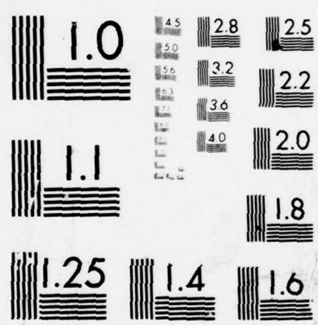
AD A068146

END

DATE FILMED

6 --79

DDC


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
5.6

TEXAS A&M UNIVERSITY
COLLEGE STATION, TEXAS 77843

INSTITUTE OF STATISTICS
Phone 713 - 845-3141

AN ANALYSIS OF THE WEIBULL FAMILY OF DISTRIBUTIONS FROM A QUANTILE BOX-PLOT PERSPECTIVE

By J. Michael White

Institute of Statistics, Texas A&M University

Technical Report No. A-3

February 1979

Texas A & M Research Foundation
Project No. 3861

"Maximum Robust Likelihood Estimation and Non-parametric Statistical Data Modeling"

Sponsored by the U.S. Army Research Office

Professor Emanuel Parzen, Principal Investigator

Approved for public release; distribution unlimited.

12 18 ARO 19 16228.4-M

Unclassified
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER	Technical Report No. A-3
2. GOVT ACCESSION NO.	
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle)	6 An Analysis of the Weibull Family of Distributions from a Quantile Box-Plot Perspective
5. TYPE OF REPORT & PERIOD COVERED	9 Technical rept.
6. PERFORMING ORG. REPORT NUMBER	
7. AUTHOR(s)	10 J. Michael White
8. CONTRACT OR GRANT NUMBER(s)	15 DAAG29-78-G-0180
9. PERFORMING ORGANIZATION NAME AND ADDRESS	Texas A&M University, Institute of Statistics, College Station, TX 77843
10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS	
11. CONTROLLING OFFICE NAME AND ADDRESS	Army Research Office, Research Triangle Park, NC 27709
12. REPORT DATE	11 Feb 1979
13. NUMBER OF PAGES	28
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	12 16p.
15. SECURITY CLASS. (of this report)	Unclassified
15a. DECLASSIFICATION DOWNGRADING SCHEDULE	
16. DISTRIBUTION STATEMENT (of this Report)	Approved for public release; distribution unlimited.
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)	NA 14 TR-A-3
18. SUPPLEMENTARY NOTES	The findings in this report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.
19. KEY WORDS (Continue on reverse side if necessary and identify by block number)	Box-plots, quantile functions, Weibull distribution, statistical data modelling

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The Weibull family of distributions has been the topic of a number of papers that attempt to define its properties and investigate techniques to detect data distributed as Weibull. This paper discusses the properties of the one parameter Weibull family, with distribution function, $F(x, \beta) = 1 - \exp(-x^{1/\beta})$, $x > 0$, for a range of values of β from a Quantile Box Plot perspective. The values of used are [1., .667, .5, .333, .25,

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

.2, .167, .125, .1]. Section 1 of the paper discusses the features of Quantile Box-Plots and the associated diagnostics. Section 2 presents the results of analyzing the Weibull family using the Quantile Box-Plot technique applied to true quantile functions. The results of simulation studies using samples from the Weibull family are presented in Section 3. Comments and conclusions are stated in Section 4.

ACCESSION for	
NTIS	White Section ☑
DDC	Buff Section ☐
UNANNOUNCED	☐
JUSTIFICATION	
BY	
DISTRIBUTION/AVAILABILITY CODES	
Dist.	SPECIAL
A	

S N 0102-LF-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

AN ANALYSIS OF THE WEIBULL FAMILY OF DISTRIBUTIONS FROM A QUANTILE BOX-PLOT PERSPECTIVE

By J. Michael White

Abstract

The Weibull family of distributions has been the topic of a number of papers that attempt to define its properties and investigate techniques to detect data distributed as Weibull. This paper discusses the properties of the one parameter Weibull family, with distribution function, $F(x, \beta) = 1 - \exp\left(-x^{1/\beta}\right),\ x > 0$, for a range of values of β from a Quantile Box Plot perspective. The values of β used are [1., .667, .5, .333, .25, .2, .167, .125, .1]. Section 1 of the paper discusses the features of Quantile Box-Plots and the associated diagnostics. Section 2 presents the results of analyzing the Weibull family using the Quantile Box-Plot technique applied to true quantile functions. The results of simulation studies using samples from the Weibull family are presented in Section 3. Comments and conclusions are stated in Section 4.

1. <u>Quantile Box-Plots and Diagnostic Measures</u>

One approach used to display and graphically analyse a batch of data is the Box-Plot introduced by Tukey (1977), as modified by Parzen (1978) under the name Quantile-Box-Plot. Five values from a set of data are usually used to construct the plots -- the extremes, the upper and lower quartiles (called hinges or H values), and the median (M-value). The basic configuration of Tukey's Box-Plot display is a box of arbitrary width and of length HH equal to the upper H minus the lower H (H-spread) with a solid horizontal line drawn within the box passing through the median. The distance from the median to the lower H is called MH . Dashed lines extend vertically from the H-values connecting the H-values with the extremes. When the Box-Plot is superimposed on the sample quantile function, $\tilde{Q}(u)$, $0 \le u \le 1$, it is called a Quantile Box-Plot.

The quantile function, $Q(u) = F^{-1}(u)$, $0 \le u \le 1$ of a given distribution function F can be estimated for a batch of data, $X_1, \ldots, X_n$, by

$$\tilde{Q}(u) = \tilde{F}^{-1}(u) = \inf[x : \tilde{F}(x) \ge u] ,$$

where $\tilde{F}(x)$ is the sample distribution function; $\tilde{Q}(u)$ can be computed in terms of the order statistics $X_{(1)} \le X_{(2)} \le \ldots \le X_{(n)}$ of $X_1, X_2, \ldots, X_n$ by

$$\tilde{Q}(u) = X_{(j)} , \quad \frac{j-1}{n} < u \le \frac{j}{n} .$$

The above definition of $\tilde{Q}(u)$ gives a piecewise constant function. It is more convenient to define $\tilde{Q}(u)$ as a piecewise linear function and divide the interval $(0, 1)$ into $2n$ subintervals. Define

$$\tilde{Q}\left(\frac{2j-1}{2n}\right) = X_{(j)} , \quad j = 1, 2, \ldots, n .$$

For $u \in [\frac{2j-1}{2n}, \frac{2j+1}{2n}]$, $j = 1, 2, \ldots, n-1$, define $\tilde{Q}(u)$ using linear interpolation by

$$\tilde{Q}(u) = n\left(u - \frac{2j-1}{2n}\right) X_{(j+1)} + n\left(\frac{2j+1}{2n} - u\right) X_{(j)} .$$

In particular, $\tilde{Q}(\frac{j}{n}) = \frac{1}{2}(X_{(j+1)} + X_{(j)})$. This definition has the merit that $\tilde{Q}(.5)$ equals the sample median as usually defined:

$$\tilde{Q}(.5) = X_{(m+1)} \quad \text{if } n = 2m+1 \text{ is odd ,}$$

$$= \frac{1}{2}(X_{(m)} + X_{(m+1)}) \quad \text{if } n = 2m \text{ is even.}$$

The Quantile Box-Plot obtained by superimposing the Box-Plot on $\tilde{Q}(\cdot)$ consists of a box of width extending from .25 to .75 and

79 04 26 032

length extending from $\tilde{Q}(.25)$ to $\tilde{Q}(.75)$. A horizontal line is drawn through $\tilde{Q}(.5)$ from .25 to .75 . This box so drawn is more aptly labeled the H-box. An E-box can also be drawn using the upper eighth $= \tilde{Q}(.875)$ (upper E-value) and the lower eighth $= \tilde{Q}(.125)$ (lower E-value) as the upper and lower bounds; .125 and .875 are the left and right bounds of the box. A D-box can be drawn as well using the upper sixteenth $= \tilde{Q}(.9375)$ (upper D-value) and the lower sixteenth $= \tilde{Q}(.0625)$ (lower D-value) as the upper and lower bounds; .0625 and .9375 are the left and right bounds of the box.

An approximate 95% confidence interval for the population median can be drawn using a vertical line of length $HH/\sqrt{n}$ centered on $\tilde{Q}(.5)$ (see McGill, Tukey, and Larsen (1978) and Parzen (1978)).

$\tilde{Q}(u)$ is useful for detecting the presence of outliers, modes, and the existence of two populations. Flat slots in $\tilde{Q}(u)$ indicate modes. Sharp rises in $\tilde{Q}(u)$ for u near 0 or 1 suggest outliers; sharp rises in $\tilde{Q}(u)$ within the boxes lead one to suspect the existence of two (or more) populations. The location of the box is useful for detecting symmetry and long or short tailedness. One can readily spot a symmetric batch of data if the median is centered within each of the boxes. One can detect data from a long tailed or a short tailed population by comparing the ratio of the lengths of the boxes, using as ideal values the ratios obtained from a normal population.

While the Quantile Box-Plot is a useful graphical tool for exploratory data analysis, accompanying diagnostics provide analytical tools to help one to interpret the plots more objectively. A seven point

summary provided by the M, H, E and D values of a batch of data is the basis for the following diagnostic measures.

The five mid-summaries of interest are of the form

$$\tilde{\mu}(p) = \frac{1}{2}\left[\tilde{Q}(1-p) + \tilde{Q}(p)\right] \text{ for } p = .5, .25, .125, .0625 ,$$

giving us

Med $= \tilde{\mu}(.5)$ = sample median.

Mid H $= \tilde{\mu}(.25) = 1/2$ (upper H lower H)

Mid E $= \tilde{\mu}(.125) = 1/2$ upper E + lower E)

Mid E $= \tilde{\mu}(.0625) = 1/2$ (upper D + lower D)

$\tilde{\mu}(0) = 1/2$ (upper extreme + lower extreme)

= mid extreme

$\overline{X} = \Sigma X_{i/n}$ is computed as well.

A "quick and dirty" estimator of μ in the model H_0: $Q(u) = \mu + \sigma Q_0(u)$, when $Q_0(u)$ is symmetric in the sense that $Q_0(u) = -Q_0(1-u)$, is provided by $\tilde{\mu}^* = (1/4)$(Med + Mid H + Mid E + Mid D) .

The scale summaries of interest are of the form

$$\tilde{\sigma}(p) = \frac{\tilde{Q}(1-p) - \tilde{Q}(p)}{Q_0(1-p) - Q_0(p)} \quad \text{for } p = .25, .125, .0625 .$$

When H_0 holds, $\tilde{\sigma}(p)$ is an approximately unbiased estimator for σ .

Two basic quantile functions Q_0 are $Q_0 = \Phi^{-1}$ (normal) and $Q_0 = -\log(1-u)$ (exponential). The resulting scale summaries are:

$HH/HHnor = \tilde{\sigma}_{\Phi}(.25)$ $\quad$ $HH/HHexp = \tilde{\sigma}_{exp}(.25)$

$EE/EEnor = \tilde{\sigma}_{\Phi}(.125)$ $\quad$ $EE/EEexp = \tilde{\sigma}_{exp}(.125)$

$DD/DDnor = \tilde{\sigma}_{\Phi}(.0625)$ $\quad$ $DD/DDexp = \tilde{\sigma}_{exp}(.0625)$

A quick and dirty estimator of σ under H_0 is provided by $\tilde{\sigma}^* = 1/ [\tilde{\sigma}(.25) + 2\tilde{\sigma}(.125) + 2\tilde{\sigma}(.0625)]$ giving us the two estimators $\tilde{\sigma}^*_{\Phi}$ and $\tilde{\sigma}^*_{exp}$. We also compute $s = \sqrt{\sum(X_i - \overline{X})^2/(n-1)}$.

The skewness diagnostics are of the form

$$skew(p) = [\tilde{Q}.5 - \tilde{Q}(p)] / [\tilde{Q}(1-p) - \tilde{Q}(p)] \quad \text{for} \quad p = .25, .125, .0625$$

giving us

$MH/HH = skew(.25)$

$ME/EE = skew(.125)$

$MD/DD = skew(.0625)$

For data from a symmetric population, we expect $skew(p)$ to be close to .5 .

The tail diagnostics computed are of the form

$$Tail(p,q) = \log[\tilde{Q}(1-p) - \tilde{Q}(p)]/[\tilde{Q}(1-q) - \tilde{Q}(q)]$$

$$\text{for} \quad p,q = .25, .125, .0625 .$$

$\log(HH/EE) = Tail(.25, .125)$

$\log(EE/DD) = Tail(.125, .0625)$

These values are to be compared with

$$Tail_0(p,q) = \log[Q_0(1-p) - Q_0(p)]/[Q_0(1-q) - Q_0(q)]$$

and $\quad Tail_{\Phi}(p,q) = \log[\Phi^{-1}(p)/\Phi^{-1}(q)]$

Using the values obtained for the normal distribution as the ideal situation, if a batch of data passes the test for symmetry, it is checked for normality by comparing $Tail(p,q)$ with $Tail_{\Phi}(p,q)$. $Tail(p,q)$ significantly larger than $Tail_{\Phi}(p,q)$ indicates a long tailed distribution while $Tail(p,q)$ smaller than $Tail_{\Phi}(p,q)$ indicates a short tailed (e.g. uniform) or bimodal distribution.

The true values which we compare the sample values to are:

$Tail_{\Phi}(.25, .125) = -.5339$

$Tail_{\Phi}(.125, .0625) = -.2879$

$Tail_{exp}(.25, .125) = -.5717$

$Tail_{exp}(.125, .0625) = -.3305$

2. Quantile Box-Plots of Theoretical Weibull Distributions

Our purpose in undertaking this study is twofold. First we desired to discover what the Quantile Box-Plots would look like and what values of the diagnostics we would obtain for the Weibull family of distributions. We also wanted to discover for which values of the Beta parameter we

might expect data from a Weibull distribution to look like data from more standard distributions, in particular, the Normal and Exponential distributions.

The values of the diagnostics for the family of Weibull distributions analyzed is summarized in Table I. Figures 1 to 9 display the Quantile Box-Plots for the distributions analyzed. The results were obtained by evaluating $Q_\beta(u) = (-\log(1-u))^\beta$ for $u = \frac{j}{200}$, $j = 1,\ldots,199$. Thus there is slight approximation error for some statistics.

When analyzing the diagnostics, bear in mind that one of our purposes is to discover values that conform to what we would expect for the normal and/or exponential distributions. Hence, in analyzing the midsummaries we are not only interested in trends in the diagnostics but also in comparing the values to the accepted location diagnostic for normal data, $\overline{X}$. For $\beta = 1$, we notice a great trend upwards in the value of the midsummaries indicating sharp skewness in the distribution. The mid-summary average $\tilde{\mu}^*$ is much greater than $\overline{X}$. For $\beta = .667$ and $.5$ this trend becomes more gradual as the distribution loses its skewness and $\tilde{\mu}^*$ approaches $\overline{X}$. For $\beta = .333$ the upward trend is slight and $\tilde{\mu}^*$ has converged to $\overline{X}$ reflecting the symmetry of the distribution.

We see that a gradual downward trend has developed for $\beta = .25$ and $.2$ but $\tilde{\mu}^*$ still approximately equals $\overline{X}$. The downward trend continues for $\beta = .167$, $.125$ and $.1$ with $\tilde{\mu}^*$ becoming slightly smaller

than $\overline{X}$. Remember that $Q(u)$ takes on only positive values so the progressively larger left tail as $\beta \to 0$ is balanced by a few large values in computing $\overline{X}$.

For the scale summaries we are interested in detecting trends in the diagnostics and also in comparing $\tilde{\sigma}^*_\Phi$ to s = standard error and $\sigma_0 = \int_0^1 f_0Q_0(u)q(u)du$ where $f_0Q_0(u) = \phi(\Phi^{-1}(u))$. For a discussion of the use of σ_0 as an estimate of scale see Parzen (1977), (1978).

Of course for $\beta = 1$ the Weibull distribution is the same as the Exponential $(\lambda = 1)$ and so, as expected, the ratio of the $\tilde{\sigma}(p)$ to the expected exponential value is 1. Notice the influence of the approximation error in computing DD/DDexp. We notice an upward trend in the values of the scale summaries when compared to the expected normal values. The value of s is larger than σ_0 which is above the range of $\tilde{\sigma}^*_\Phi$.

For $\beta = .667$ the scale summaries for the normal case are also approximately equal and $\tilde{\sigma}^*_\Phi$ is slightly less than σ_0 but considerably less than s. For the exponential case, there is a downward trend with $\tilde{\sigma}^*_{exp}$ larger than σ_0.

For $\beta = .5$ and $.333$, there is a downward trend in the values of $\tilde{\sigma}$ for both the normal and exponential case with $\tilde{\sigma}^*_\Phi$ in the same range but slightly larger than σ_0 or s; $\tilde{\sigma}^*_{exp}$ is larger than σ_0

in both cases. There is no trend in the values of the scale summaries for the normal case for $\beta = .25$ and $.2$ indicating symmetry; $\tilde{\sigma}^*_\phi$ approximately equals σ_0 and s .

In analyzing the skewness measures one can see a well defined pattern for the Weibull distribution. A symmetric distribution forces the values of the diagnostics to equal .5 . A large right tail yields small values of the skewness measures and a large left tail yields values close to 1 . For $\beta \geq .333$, there is a downward trend in the values of the diagnostics with very small values for $\beta = 1$ and values close to .5 with only slight trending for $\beta = .333$. For $\beta \leq .25$ there is an upward trend in the values of the skewness measures with a general tendency to get larger values and a steeper trend as β gets smaller. This reflects the same information that one obtains from the Quantile Box-Plots.

The tail measure diagnostics are significant when interpreted in conjunction with the skewness diagnostics. If we have symmetric data we should be interested in how the tails compare to the tails of a normal distribution. For $\beta = 1$, the values of the tail measures are close to what we would expect from the exponential distribution. For $\beta = .667$ log (HH/EE) is far from the exponential value and close to the normal value; log (EE/DD) is closer to the normal value also. Likewise, for $\beta = .5$ the values are closer to the normal values than to the exponential values. For $\beta \leq .333$ there is a general trend downward

in the value of log (HH/EE) and log (EE/DD) . Values of log (HH/EE) are close to the normal and far from the exponential for $\beta = .333, .25, .2$ and $.167$; values of log (EE/DD) are between the normal and exponential values but closer to the normal values. For $\beta = .125$ and $.1$, the values of both diagnostics are between the normal values and exponential values with log (HH/EE) closer to the normal values.

The Quantile Box-Plots tell the same story as the diagnostics. Q(u) is skewed and has a large upper tail for $\beta = 1$. This trend is modified progressively for $\beta = .667$ and $.5$. Q(u) is symmetric for $\beta = .333$ and $.25$ and looks very much like $\Phi^{-1}(u)$. For $\beta \leq .2$ we see the development of a longer lower tail with the tail very pronounced for $\beta = .1$. The Quantile Box Plot for $\beta = .1$ looks almost the opposite as that for $\beta = .667$

Table I

Values of Diagnostic Measures for Theoretical Weibull Distributions

	1	.667	.500	.333	.250	.200	.167	.125	.100
7-Point Summary									
Upper D	2.817	1.994	1.678	1.412	1.295	1.230	1.188	1.138	1.109
Upper E	2.079	1.629	1.442	1.276	1.201	1.158	1.130	1.096	1.076
Upper H	1.386	1.243	1.177	1.115	1.085	1.068	1.056	1.042	1.033
Median	.693	.783	.833	.885	.912	.929	.941	.955	.964
Lower H	.288	.436	.536	.660	.732	.779	.812	.856	.883
Lower E	.134	.261	.365	.511	.604	.669	.715	.777	.818
Lower D	.062	.156	.249	.395	.498	.573	.629	.706	.757
Mid-Summary									
$\tilde{C}_M$ Median	.693	.783	.833	.885	.912	.929	.941	.955	.964
$\tilde{C}_H$ = Mid H	.837	.840	.857	.888	.909	.923	.934	.949	.958
$\tilde{C}_E$ = Mid E	1.106	.954	.904	.894	.903	.913	.922	.937	.947
$\tilde{C}_D$ = Mid D	1.439	1.075	.963	.904	.897	.901	.908	.922	.933
$\tilde{C}^*$ = Average	1.019	.911	.889	.893	.905	.917	.926	.941	.950
$\overline{X}$ = X Bar	.997	.902	.886	.893	.906	.918	.928	.942	.951
$HH/\sqrt{n}$	.110	.081	.064	.045	.035	.029	.024	.019	.015
Scale Summary									
HH/HHnor	.815	.599	.475	.337	.262	.214	.180	.138	.111
EE/EEnor	.846	.595	.468	.333	.259	.213	.180	.138	.112
DD/DDnor	.898	.599	.466	.331	.260	.214	.182	.141	.115
$\tilde{\sigma}_*$ = Average	.853	.597	.470	.334	.260	.213	.181	.139	.113
HH/HHexp	1.000	.735	.584	.414	.321	.262	.222	.169	.137
EE/EEexp	1.000	.703	.553	.393	.306	.251	.213	.164	.133
DD/DDexp	1.017	.679	.528	.376	.294	.243	.207	.160	.130
$\tilde{\sigma}_{exp}$ = Average	1.006	.706	.555	.394	.307	.252	.214	.164	.133
S. D.	.985	.611	.463	.325	.255	.210	.180	.140	.114
σ_0	.9225	.6004	..4625	.3268	.2557	.2108	.1796	.1388	.1132
Skewness									
MH/HH	.3690	.4302	.4620	.4943	.5105	.5203	.5268	.5349	.5398
ME/EE	.2876	.3816	.4339	.4886	.5164	.5331	.5443	.5583	.5666
MD/DD	.2291	.3411	.4083	.4816	.5195	.5424	.5577	.5767	.5881
Tail									
Log(HH/EE)	-.5715	-.5270	-.5184	-.5201	-.5250	-.5294	-.5328	-.5378	-.5412
Log(EE/DD)	-.3477	-.2954	-.2836	-.2843	-.2900	-.2953	-.2995	-.3058	-.3100

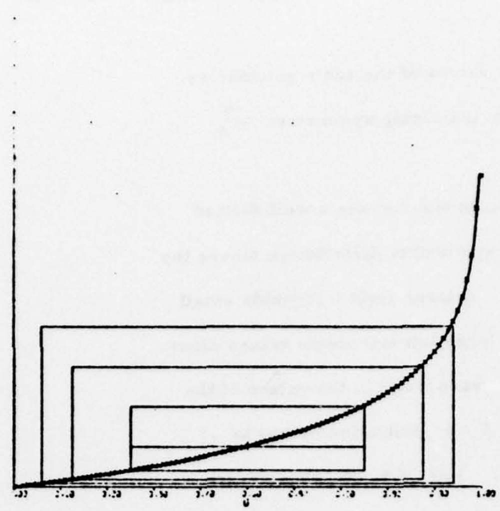


figure 1

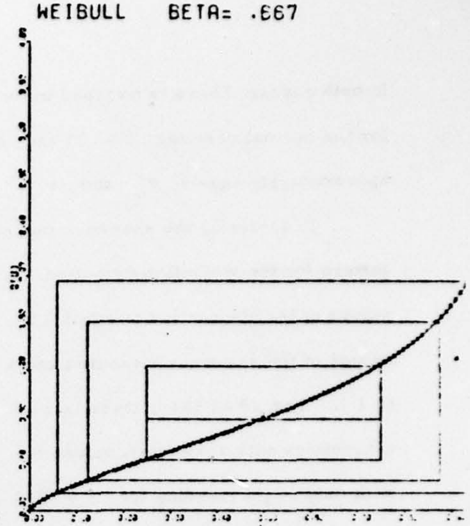


figure 2

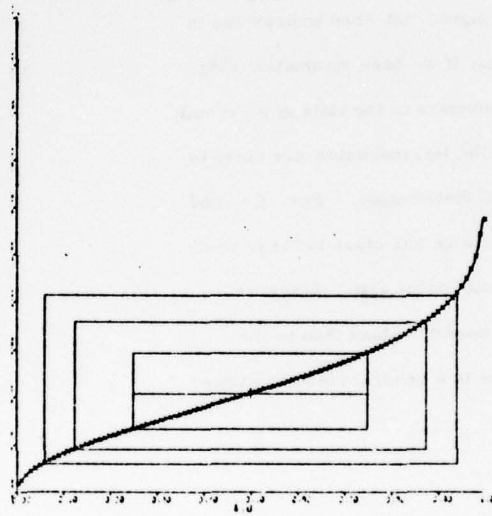


figure 3

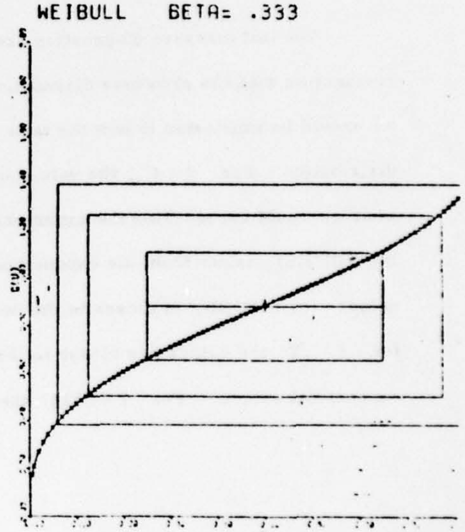


figure 4

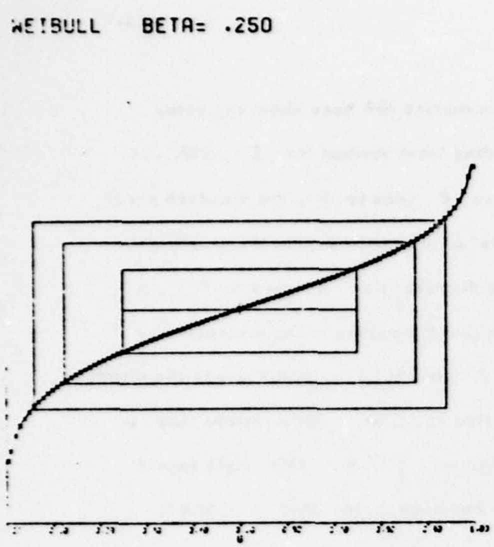


figure 5

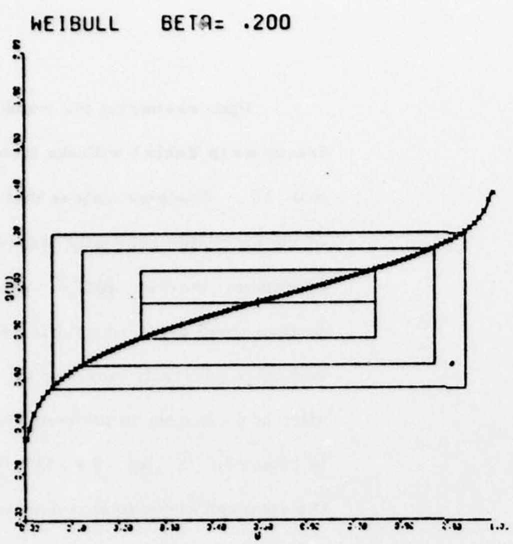


figure 6

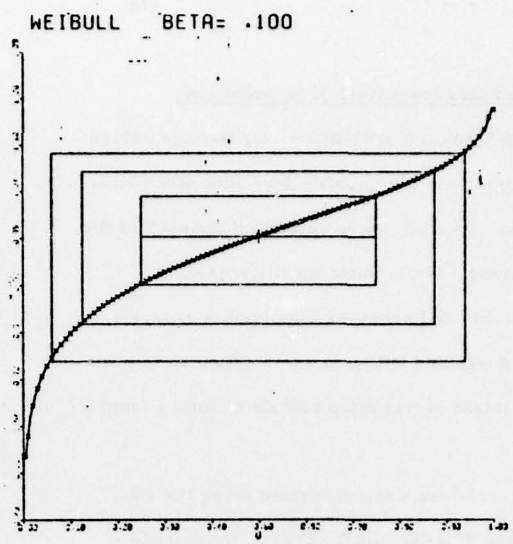


figure 9

figure 7

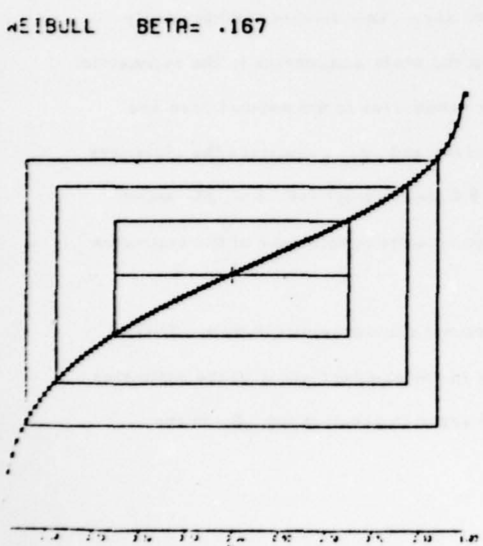
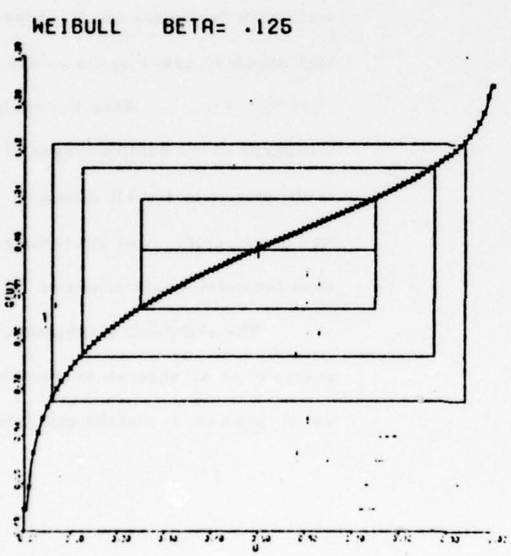


figure 8



3. Quantile Box-Plots of Samples from Weibull Distributions

Simulation samples from Weibull distributions, for various values of the β-parameter, were analyzed via the Quantile Box-Plot procedure. Six simulations with sample size $n = 100$ were conducted for each of the nine values of β used in the study. While more simulations of different sample sizes are desirable and necessary for further analysis, it is hoped that the simulations conducted will give some notion as to what kind of values and deviation we might expect from real data coming from a Weibull distribution.

Uniform (0,1) random variables were generated using the subroutine RANDU, available on the TIMESBOARD library; these were transformed by $Y = (-\log(1-u))^{\beta}$ giving us our simulated data from the Weibull distribution. The diagnostics were computed and Quantile Box-Plots were drawn for each of the samples. We then computed the sample mean and standard error of each diagnostic using the six estimates thus obtained. An approximate 95% confidence interval for the true value, μ, of the diagnostic could be formed using $\overline{X} \pm 1.05s$ where $\overline{X}$ is the sample mean of our estimates and s is the standard error of our estimates with $\sqrt{n}(\overline{X} - \mu)/s$ approximately distributed as a Student's t with 5 d.f. The results of these computations are presented in Table II. The first number in each entry in the table is $\overline{X}$ followed by s.

Upon examining the midsummaries one sees about the same trends as in Table I with the trending least evident for $\beta = .337, .25$ and $.20$. One also notices that as β goes to 0, the standard error of the estimates generally decreases. One notices that the length of the confidence interval $HH/\sqrt{n}$ also decreases as β goes to 0. A further trend evident in Table I is that the values of the midsummary statistics generally increase as β goes to 0 probably due to the elimination of a constant in the computation of $Q(u)$. One notices that $\tilde{\mu}^*$ is closer to $\overline{X}$ for $\beta \leq .333$ than for $\beta = .5$. One might expect the standard error to increase as one goes from Med to Mid D but this is not evident.

The scale summary diagnostics reveal the same kind of trending in the standard error of the estimates. The standard error of the estimates decreases as β goes to zero. One detects a surprisingly high standard error in the values of the scale summaries in the exponential case for $\beta = 1$. When the scale summaries in the normal case are compared to the sample standard error and σ_Q, one sees the closeness of the statistics for all values of $\beta \leq .667$ except for $\beta = .25$ which was unexpected. It is interesting to note the consistency of the estimates even for quite small values of β.

The skewness diagnostics reveal one surprising result. There seems to be an absence of trending in the standard error of the estimates as β goes to 0 and the standard error is greatest for β in the

range (.667, .167) which includes the normal case. One also sees a wider range of values of the diagnostic than one expects when the results in Table II are compared to those of Table I. As expected, one finds values of the diagnostic closest to .5 for β in the range (.333, .167). However, the lack of symmetry in the tails of the Weibull distribution with $\beta = .167$ is more obvious when using MD/DD.

Upon examining the tail diagnostic, one again notices no trending in the standard error of the estimates as β goes to 0. One would probably not use this diagnostic for $\beta = 1$ and it is not surprising that the greatest standard error is for $\beta = 1$. The trending in the values of the diagnostic that was evident in Table I is also absent. The values for $\beta = 1$ are not close to the expected exponential values of -.5717 and -.3305. However one does notice values of -.5339 and -.2879 for β in the range (.333, .20). It is curious that for $\beta = .1$ we also obtain values close to the expected normal values but one would disregard these values, having already disregarded normality based on the skewness diagnostics. The values of the diagnostics in Table II show wide deviations from their true value as found in Table I and one sees little trending in the over or under-estimations.

A selected number of Quantile Box-Plots are included in the Appendix to give some idea as to what kind of deviations from the ideal one can expect.

Table II

	1	.667	.500	.333	.250
Median	.675/.485	.769/.097	.806/.061	.885/.020	.910/.030
Mid H	.854/.065	.846/.055	.845/.070	.894/.029	.915/.033
Mid E	1.085/.132	.945/.091	.899/.037	.896/.036	.921/.029
Mid D	1.514/.155	1.055/.106	.947/.025	.908/.069	.908/.039
$\tilde{\mu}^*$ = Average	1.032/.082	.903/.073	.874/.045	.896/.036	.914/.030
X	1.011/.074	.892/.082	.869/.044	.897.036	.912/.029
HH/$\sqrt{n}$	.116/.017	.082/.011	.064/.012	.046/.005	.034/.004
HH/HHnor	.858/.124	.605/.078	.477/.084	.340/.039	.249/.026
EE/EEnor	.847/.093	.601/.080	.461/.058	.346/.032	.251/.021
DD/DDnor	.946/.090	.589/.070	.455/.038	.352/.035	.259/.020
$\tilde{\sigma}^*_\phi$ = Average	.884/.065	.598/.062	.465/.057	.346/.029	.253/.020
HH/HHexp	1.053/.152	.743/.096	.586/.104	.418/.048	.306/.032
EE/EEexp	1.002/.109	.710/.094	.545/.069	.409/.038	.298/.025
DD/DDexp	1.072/.102	.668/.080	.516/.043	.398/.040	.293/.023
$\tilde{\sigma}^*_{exp}$ = Average	1.042/.078	.707/.082	.549/.068	.409/.035	.299/.023
S. E.	1.029/.079	.595/.079	.461/.042	.337/.028	.246/.018
σ_0					
MH/HH	.3459/.0578	.3995/.1047	.4397/.0829	.4783/.0346	.4847/.0609
ME/EE	.2900/.0196	.3754/.0743	.4072/.0465	.4876/.0229	.4825/.0476
MD/DD	.2127/.0332	.3433/.0563	.3963/.0415	.4812/.0471	.5041/.0417
log(HH/EE)	-.5255/.1563	-.5264/.1129	-.5056/.1011	-.5538/.0970	-.5451/.0758
log(EE/DD)	-.3994/.1609	-.2697/.0321	-.2793/.0487	-.3024/.0660	-.3160/.0497

Table II (continued)

	.200	.167	.125	.100
Median	.937/.020	.947/.032	.966/.014	.965/.010
Mid H	.939/.014	.943/.015	.958/.019	.949/.011
Mid E	.931/.016	.936/.021	.945/.021	.937/.013
Mid D	.918/.022	.920/.022	.932/.023	.923/.021
$\tilde{\mu}^*$ = Average	.931/.012	.937/.018	.950/.018	.943/.013
$\overline{X}$	.932/.013	.938/.016	.951/.016	.946/.011
HH/$\sqrt{n}$	.028/.022	.026/.005	.018/.022	.015/.001
HH/HHnor	.210/.018	.192/.038	.134/.015	.111/.008
EE/EEnor	.214/.020	.187/.018	.131/.013	.113/.012
DD/DDnor	.215/.026	.182/.015	.136/.005	.114/.010
$\tilde{\sigma}^*$ = Average	.213/.017	.187/.022	.134/.009	.113/.009
HH/HHexp	.258/.022	.236/.047	.164/.019	.136/.010
EE/EEexp	.254/.023	.221/.022	.154/.016	.134/.014
DD/DDexp	.243/.029	.207/.017	.154/.006	.130/.012
$\tilde{\sigma}^*_{exp}$ = Average	.252/.019	.221/.026	.158/.011	.133/.011
S.E.	.213/.017	.182/.018	.135/.008	.115/.009
σ_0				
MH/HH	.5023/.0492	.5088/.0940	.5462/.0629	.6092/.0382
ME/EE	.5130/.0512	.5252/.0497	.5738/.0335	.6043/.0271
MD/DD	.5270/.0562	.5465/.0691	.5827/.0385	.6205/.0391
log(HH/EE)	-.5542/.1025	-.5182/.1443	-.5073/.0758	-.5536/.0635
log(EE/DD)	-.2861/.1010	-.2656/.0909	-.3316/.1172	-.2973/.0645

4. Comments and Conclusions

This study is lacking in one outstanding aspect; more simulations are needed for different sample sizes. The distribution theory of the skewness and tail diagnostics can be computed with difficulty using the distribution of functions of order statistics but more work in this direction is also needed. One should be cautious about drawing conclusions from a limited Monte Carlo study. However, it seems that one can draw some conclusions from this analysis of the Weibull distribution.

First one notices a surprising predictability in the values of the midsummary and scale summary diagnostics for rather small values of β. It appears that in the simulations $\mu(p)$ is closer to its true value than $\overline{X}$ to its true value for almost the entire range of β values analyzed. The scale summary diagnostics are not quite so predictable but are in general at least as consistent as the s statistic. It seems that the skewness diagnostics are suspectible to be influenced by mild aberations in the data. The same seems to be true of the tail diagnostics. This could account somewhat for the unexpected fluctuations in the value of the statistics.

While one can very easily detect from looking at the diagnostics obtained from the true quantile function that the Weibull distribution with $\beta = 1$ is the same as the Exponential ($\lambda = 1$) distribution, it is not so obvious to detect a batch of data from the Weibull ($\beta = 1$)

distribution. However it seems that data from a Weibull distribution with β in the range (.33, .20) could easily be classified as Normal data when analyzed from a Quantile Box-Plot perspective. The analysis of the true Quantile function supports this contention with all evidence pointing to the approximate equivalence of the two distributions for β in the range (.33, .20). It has been suggested that the Weibull distribution as β goes to zero closely resembles an extreme value distribution. The Quantile Box plots point to this equivalence. However, the extreme value distribution has not been analyzed from a Quantile Box-Plot perspective; here is another area for further exploration.

The values of the diagnostics as given in Table I have inherent worth in that they are the true values of the statistics for the Weibull distribution for the range of β values analyzed. The corresponding Quantile Box-Plots also have value in that they represent the true Quantile function with the associated boxes. Table II and the Quantile Box-Plots associated with the simulations are interesting in that they reveal the deviations from the ideal that we can expect.

References

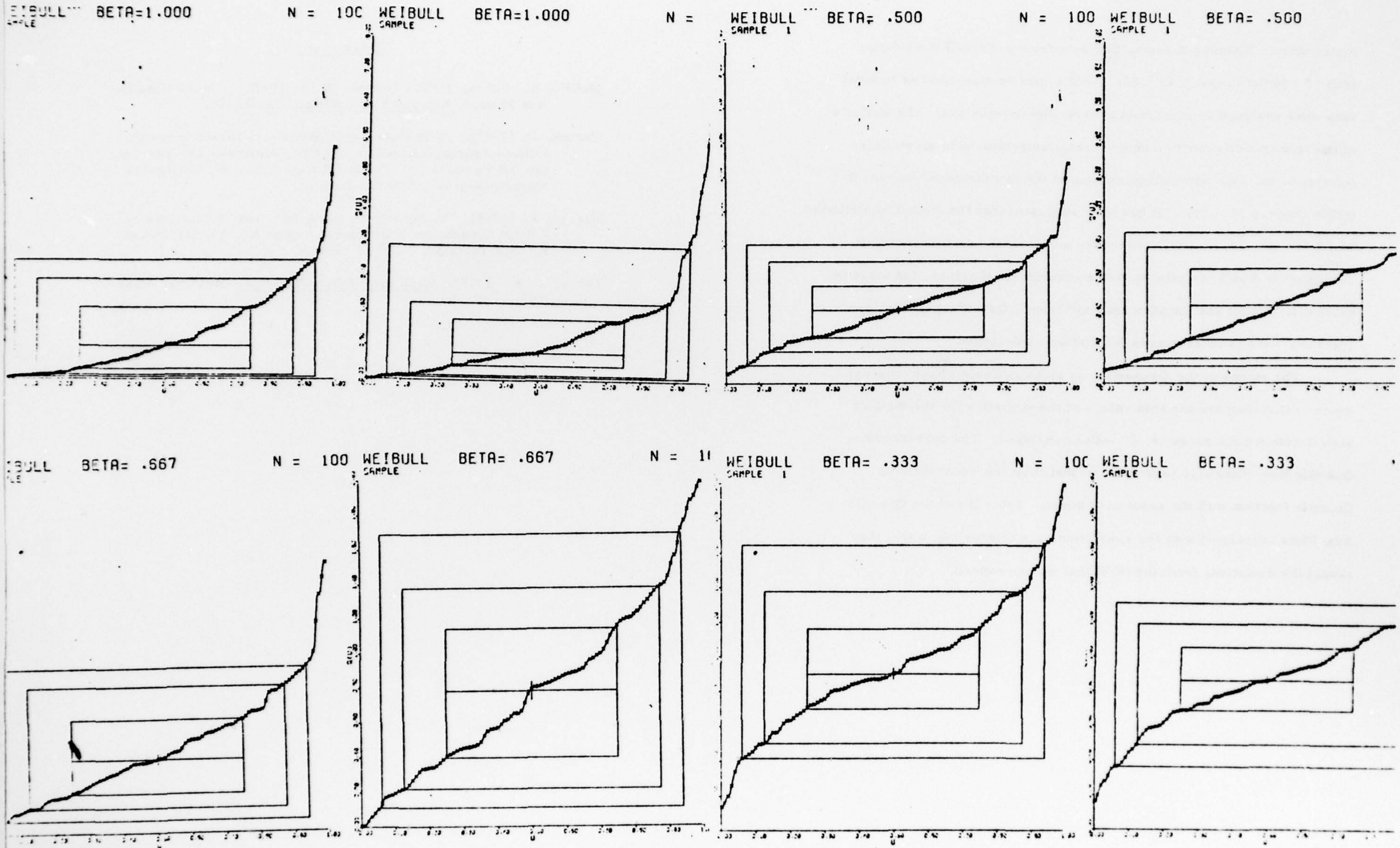
McGill, R.; Tukey, J. W.; Larsen, W. A. [1978]. "Variations of Box Plots," American Statistician, 32, 12-16.

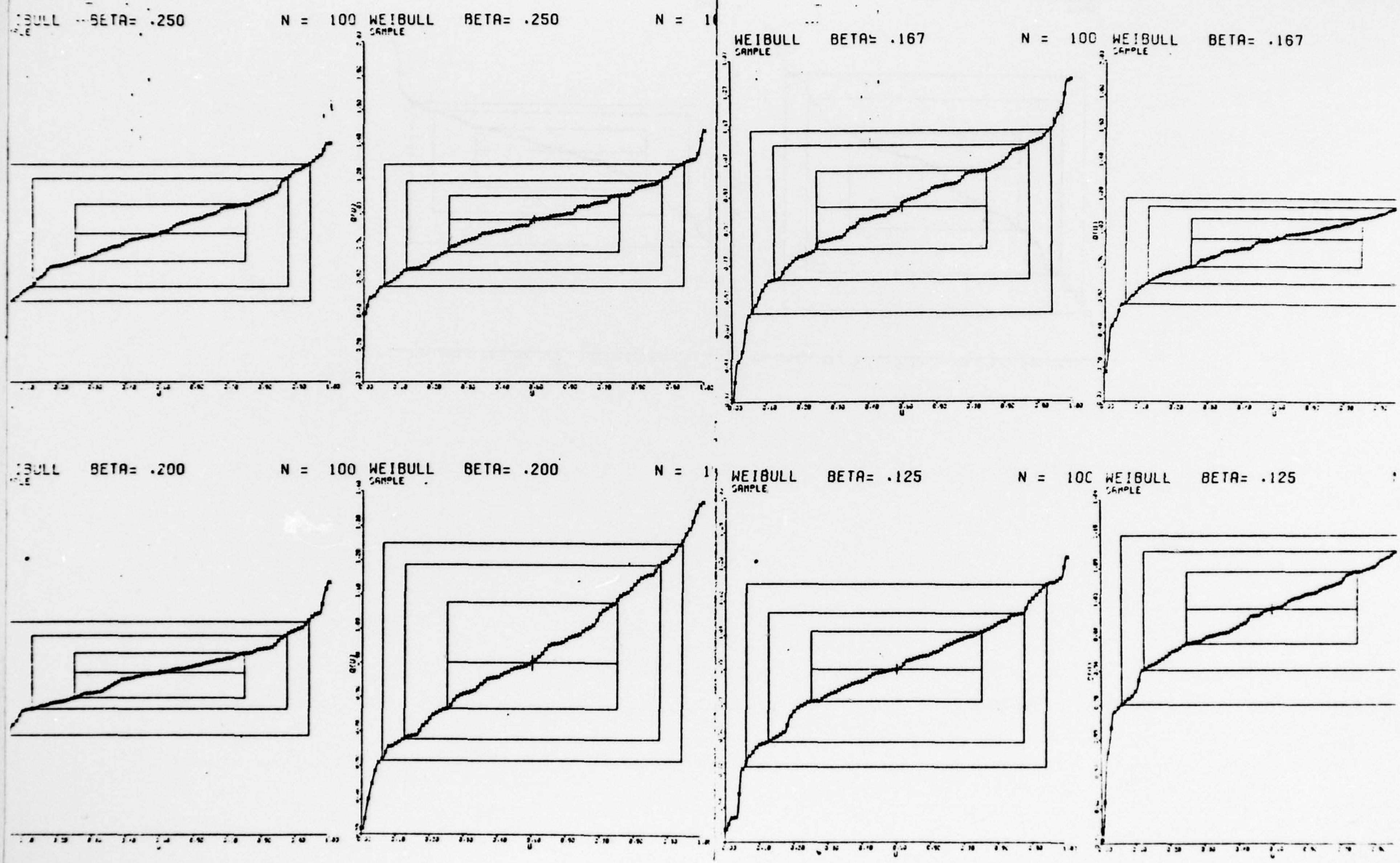
Parzen, E. [1977]. "Non Parametric Statistical Data Science (A Unified Approach Based on Density Estimation and Testing for 'White Noise')." Technical Report No. 47, Statistical Science Division, SUNY at Buffalo.

Parzen, E. [1978]. "A Density-Quantile Function Perspective on Robust Estimation." Technical Report No. 60, Statistical Science Division, SUNY at Buffalo.

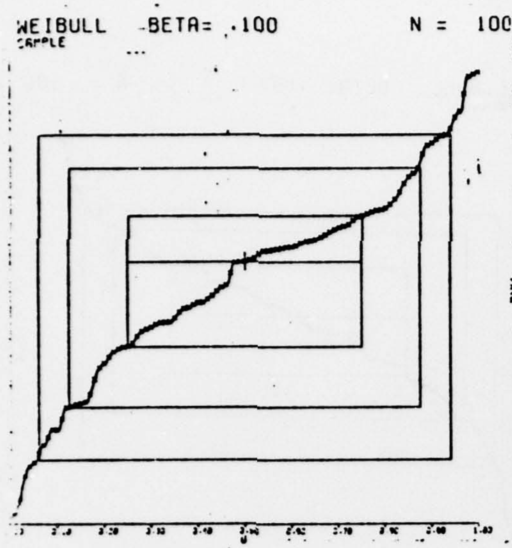
Tukey, J. W. [1977]. Exploratory Data Analysis. Reading, Mass.: Addison Wesley.

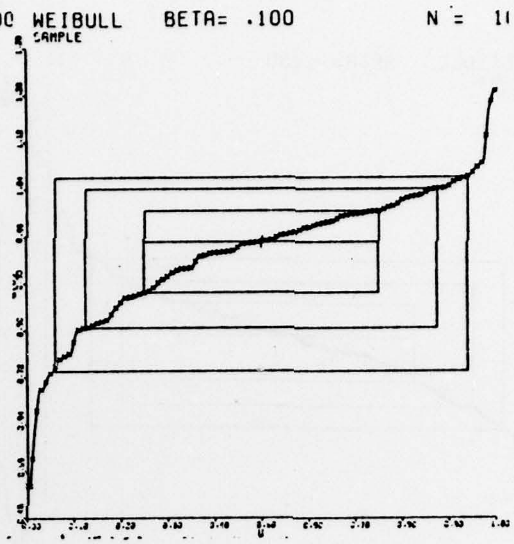
Appendix


BETA=1.000
N = 10C WEIBULL BETA=1.000
SAMPLE
N = WEIBULL BETA= .500
SAMPLE 1
N = 100 WEIBULL BETA= .500
SAMPLE 1
BETA= .667
N = 100 WEIBULL BETA= .667
SAMPLE
WEIBULL BETA= .333
SAMPLE 1
N = 10C WEIBULL BETA= .333
SAMPLE 1



IBULL BETA= .250
N = 100 WEIBULL BETA= .250
SAMPLE
N = 1
WEIBULL BETA= .167
SAMPLE
N = 100 WEIBULL BETA= .167
SAMPLE
IBULL BETA= .200
N = 100 WEIBULL BETA= .200
SAMPLE
N = 1
WEIBULL BETA= .125
SAMPLE
N = 100 WEIBULL BETA= .125
SAMPLE


WEIBULL BETA= .100 N = 100
SAMPLE


WEIBULL BETA= .100 N = 1(
SAMPLE